

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:20:34 ; Search time 17.21 Seconds  
(without alignments)  
103.206 Million cell updates/sec

Title: US-09-331-631a-25\_COPY\_31\_85  
Perfect score: 315  
Sequence: 1 ENPKHMKCLOSCNSERSDYR.....EECEGEELPRPRPQHPER 55

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	315	100.0	605	1 GLCA_SOYBN	P13916 glycine max
2	259.5	82.4	639	1 GLCX_SOYBN	P11827 glycine max
3	89	28.3	614	1 AH11_ARAHY	P43237 arachis hyp
4	89	28.3	626	1 AH12_ARAHY	P43338 arachis hyp
5	69	21.9	246	1 TOMB_PASHA	P72204 pasteurella
6	62.5	19.8	588	1 VCLB_GOSHI	P09801 gossypium h
7	62.5	19.8	3051	1 YMX3_CAEEL	P34576 caenorhabdi
8	62	19.7	380	1 EAR2_MOUSE	P43136 mus musculu
9	60	19.0	390	1 EAR2_MOUSE	P09017 rattus norv
10	60	19.0	2203	1 CCAD_RAT	P27732 rattus norv
11	59.5	18.9	174	1 BARI_CHITE	P02849 chironomus
12	59	18.7	344	1 MST1_DROHY	P08695 drosophila
13	59	18.7	713	1 TS44_GIALA	P21849 giardia lam
14	59	18.7	1391	1 MST2_DROHY	P08696 drosophila
15	58.5	18.6	638	1 OARI_LYMTST	077408 lymnaea sta
16	58.5	18.6	1610	1 CCAD_MESAU	099244 mesocricetu
17	58	18.4	81	1 TRGS_TACTR	P81281 tachyplesu
18	58	18.4	798	1 TRB1_MOUSE	P09055 mus musculu
19	58	18.4	799	1 TRB1_MOUSE	P49134 rattus norv
20	58	18.4	803	1 TRB1_MOUSE	P07228 rattus norv
21	58	18.4	2161	1 ITB1_CHICK	P01668 gallus gall
22	57.5	18.3	131	1 CCAD_HUMAN	P01668 homo sapien
23	57.5	18.3	178	1 NI2M_BOVIN	P26377 bos taurus
24	57.5	18.3	542	1 HM18_BOVIN	P02369 bos taurus
25	57.5	18.3	542	1 HM18_BOVIN	P41934 caenorhabdi
26	57	18.1	403	1 EAR2_HUMAN	P10588 homo sapien
27	57	18.1	699	1 E7S3_MANSE	008893 manduca sex
28	56.5	17.9	491	1 CGEL_MOUSE	008457 mus musculu
29	56.5	17.9	950	1 ITB1_HUMAN	P05556 homo sapien
30	56.5	17.9	950	1 Y136_HUMAN	014149 homo sapien
31	56	17.8	479	1 FIBR_RAT	P14480 rattus norv
32	56	17.8	1257	1 FLTH_CAEEL	P34268 caenorhabdi
33	55.5	17.6	205	1 BE11_FOAMY	P14353 human spuma

34	55	17.5	190	1 BCT7_SHEEP	P50415 ovis aries
35	55	17.5	230	1 YE03_SCHPO	O13800 schizosacch
36	55	17.5	376	1 HMEV_DROME	P06602 drosophila
37	55	17.5	505	1 SRK1_SPOLA	P42686 spongilla 1
38	55	17.5	506	1 SRK4_SPOLA	P42680 spongilla 1
39	55	17.5	591	1 MAOX_VITVI	P51615 vitis vinif
40	55	17.5	680	1 KALM_HUMAN	P23352 homo sapien
41	55	17.5	950	1 URBI_USTMA	P40349 ustilago ma
42	55	17.5	1125	1 TIE2_BOVIN	006807 bos taurus
43	55	17.5	1210	1 EGFR_HUMAN	P00533 homo sapien
44	54.5	17.3	446	1 MUC_CHICK	P01875 gallus gall
45	54.5	17.3	554	1 PERE_MOUSE	P10820 mus musculu

## ALIGNMENTS

RESULT	ID	GLCA_SOYBN	STANDARD	PRT	605 AA
AC	P13916				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.				
OS	Glycine max (Soybean).				
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:				
OC	Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I:				
OC	Fabales: Fabaceae: Papilionoideae: Glycine.				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=COTYLEDON;				
RA	MEDLINE=91355860; PubMed=2103438;				
RX	Sebastiani F.L., Farrel L.B., Schuler M.A., Beachy R.N.:				
RT	"Complete sequence of a cDNA of alpha subunit of soybean beta-				
RL	conglycinin.";				
CC	Plant Mol. Biol. 15:197-201(1990).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED				
CC	DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A				
CC	CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SPROUTING.				
CC	-1- SUBUNIT: THE ALPHA'-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN				
CC	VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.				
CC	-1- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND				
CC	VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVICTILIN, CONGLYCININ, ETC.)				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: X17698; CAA35691.1; -				
DR	PIR: S14681; FMSYBA.				
DR	HSSP: P50477; ICAM.				
DR	INTERPRO: IPR001113; -				
KW	PPAM; PF00546; Seedstore_7s; 1.				
DR	Seed storage protein; Signal; Glycoprotein; Multigene family.				
FT	SIGNAL	1	22		
FT	PROPEP	23	62		
FT	CHAIN	63	605		
FT	CARBOHYD	261	261		
FT	CARBOHYD	517	517		
FT	SEQUENCE	605 AA; 70293 MW; CBBBA30506BBBC57 CRC64;			

Query Match 100.0%; Score 315; DB 1; Length 605;  
Best local Similarity 100.0%; Pred. No. 1.6e-26;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ENKHNKCLQSCNSERDSTRNQCACNCLKVEKECEGEIIPRRPQHPER 55  
 DB 31 ENKHNKCLQSCNSERDSTRNQCACNCLKVEKECEGEIIPRRPQHPER 85

## RESULT 2

ID GLCX\_SOYBN STANDARD: PRT: 639 AA.  
 AC P11827;  
 DT 01-OCT-1989 (rel. 12, Created)  
 DT 01-OCT-1989 (rel. 12, last sequence update)  
 DT 01-AUG-1992 (rel. 23, last annotation update)  
 DE BETA-CONGLYCININ, ALPHA' CHAIN PRECURSOR.  
 OS CG-1.  
 GN glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Glycyne.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-86250867; PubMed-3013879;  
 RA Doyle J.J., Schuller M.A., Godette W.D., Zenger V., Beachy R.N.,  
 RA Slightom J.L.;  
 RT "The glycosylated seed storage proteins of glycine max and Phaseolus  
 vulgaris. Structural homologues of genes and proteins.";  
 RL J. Biol. Chem. 261:9228-9238(1986).  
 RN [2]  
 RP SEQUENCE OF 340-639 FROM N.A.  
 RX MEDLINE-83143289; PubMed-6897678;  
 RA Schuller M.A., Ladin B.F., Pollaco J.C., Freyer G., Beachy R.N.;  
 RT "Structural sequences are conserved in the genes coding for the  
 alpha, alpha' and beta-subunits of the soybean 7S seed storage  
 protein.";  
 RL Nucleic Acids Res. 10:8245-8261(1982).  
 CC -1- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED  
 DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A  
 CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.  
 CC -1- SUBUNIT: THE ALPHA', ALPHA'', AND BETA-SUBUNITS ASSOCIATE IN  
 VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN  
 BODIES.  
 CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
 CONVICLIN, CONGLYCININ, ETC.).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M13759; AAB01374.1; -  
 DR EMBL: J01290; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: B24810; B24810.  
 DR HSP: P02853; 2PH.  
 DR INTERPRO: IPR001113; -  
 DR PFAM: PF00546; Seedstore\_7s; 2.  
 DR Seed storage protein; Signal; Glycoprotein; Multigene family.  
 FT SIGNAL 1 22  
 FT PROPEP 23 62  
 FT CHAIN 63 639 BETA-CONGLYCININ, ALPHA' CHAIN.  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 543 543 P -> L (IN REF. 2).  
 FT CONFLICT 549 549 M -> V (IN REF. 2).  
 FT CONFLICT 608 608 S -> T (IN REF. 2).  
 FT SEQUENCE 639 AA; 74325 MW; 469BF24C79651E3F CRC64;

Query Match 82.4%; Score 259.5; DB 1; Length 639;  
 Best Local Similarity 83.9%; Pred. No. 1.3e-20;  
 Matches 47; Conservative 5; Mismatches 1; Indels 3; Gaps 2;

QY 1 ENKHNKCLQSCNSERDSTRNQCACNCLKVEKECEGEIIPRRPQHPER 55  
 DB 31 QNPSHNKCLQSCNSERDSTRNQCACNCLKVEKECEGEIIPRRPQHPER 84

## RESULT 3

ID AH12\_ARAHY STANDARD: PRT: 614 AA.  
 AC P43237;  
 DT 01-NOV-1995 (rel. 32, Created)  
 DT 01-NOV-1995 (rel. 32, last sequence update)  
 DT 01-NOV-1997 (rel. 35, last annotation update)  
 DE ALLERGEN ARA H 1, CLONE P17 (ARA H 1).  
 OS Arachis hypogaea (Peanut).  
 GN Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Arachis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. FLORUNNER;  
 RX MEDLINE-96013631; PubMed-7560062;  
 RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;  
 RT "Recombinant peanut allergen Ara h I expression and IGE binding in  
 patients with peanut hypersensitivity.";  
 RL J. Clin. Invest. 96:1715-1721(1995).  
 CC -1- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
 CONVICLIN, CONGLYCININ, ETC.).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U38853; AAA60336.1; -  
 DR HSP: P50477; ICAM.  
 DR INTERPRO: IPR001113; -  
 DR PFAM: PF00546; Seedstore\_7s; 1.  
 KW Allergen.  
 KW Allergen.  
 SQ SEQUENCE 614 AA; 70283 MW; 1DDACF217EBC5F31 CRC64;

Query Match 28.3%; Score 89; DB 1; Length 614;  
 Best Local Similarity 44.1%; Pred. No. 0.017;  
 Matches 15; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 ENKHNKCLQSCNSERDSTRNQCACNCLKVE 34  
 DB 33 ENPCACRCLOSCQOEPPDLKORACESRCYKLEYD 66  
 RESULT 4  
 AH12\_ARAHY STANDARD: PRT: 626 AA.  
 AC P43238;  
 DT 01-NOV-1995 (rel. 32, Created)  
 DT 01-NOV-1995 (rel. 32, last sequence update)  
 DT 01-NOV-1997 (rel. 35, last annotation update)  
 DE ALLERGEN ARA H 1, CLONE P41B (ARA H 1).  
 OS Arachis hypogaea (Peanut).  
 GN Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Arachis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. FLORUNNER;  
 RX MEDLINE-96013631; PubMed-7560062;  
 RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;  
 RT "Recombinant peanut allergen Ara h I expression and IGE binding in  
 patients with peanut hypersensitivity.";



DT 01-FEB-1994 (Rel. 28, last sequence update)  
DE 01-NOV-1997 (Rel. 35, last annotation update)  
DE HYPOTHETICAL PROTEIN T20G5.3 IN CHROMOSOME III (FRAGMENT).  
GN T20G5.3.  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; *Caenorhabditis*.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Berks M., Smith A.;  
RU Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 45 EGF-LIKE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
CC -----  
DR EMBL: Z30423; CAA83007.1; -.  
DR PIR: S42373; S42373.  
DR HSSP: P35555; IEMO.  
DR WORMPEP: T20G5.3; CE00478.  
DR INTERPRO: IPR000082; -.  
DR INTERPRO: IPR000152; -.  
DR INTERPRO: IPR000561; -.  
DR INTERPRO: IPR001881; -.  
DR INTERPRO: IPR002035; -.  
DR PFAM: PF00008; EGF\_31.  
DR PFAM: PF01390; SEA; 2.  
DR PFAM: PF00092; vwa; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 28.  
DR PROSITE: PS01186; EGF\_2; 5.  
DR PROSITE: PS01187; EGF\_CA; 1.  
KW Hypothetical protein; EGF-like domain; Repeat; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 2701 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2702 2722 POTENTIAL.  
FT DOMAIN 2723 3051 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN <1 44 EGF-LIKE 1 (PARTIAL).  
FT DOMAIN 46 94 EGF-LIKE 2.  
FT DOMAIN 100 144 EGF-LIKE 3.  
FT DOMAIN 145 192 EGF-LIKE 4.  
FT DOMAIN 194 245 EGF-LIKE 5.  
FT DOMAIN 247 296 EGF-LIKE 6.  
FT DOMAIN 313 354 EGF-LIKE 7.  
FT DOMAIN 355 402 EGF-LIKE 8.  
FT DOMAIN 404 452 EGF-LIKE 9.  
FT DOMAIN 454 503 EGF-LIKE 10.  
FT DOMAIN 514 699 vWFA.  
FT DOMAIN 705 750 EGF-LIKE 11.  
FT DOMAIN 750 794 EGF-LIKE 12.  
FT DOMAIN 805 846 EGF-LIKE 13.  
FT DOMAIN 847 892 EGF-LIKE 14.  
FT DOMAIN 892 940 EGF-LIKE 15.  
FT DOMAIN 942 990 EGF-LIKE 16.  
FT DOMAIN 992 1039 EGF-LIKE 17.  
FT DOMAIN 1043 1091 EGF-LIKE 18.  
FT DOMAIN 1093 1144 EGF-LIKE 19.  
FT DOMAIN 1146 1195 EGF-LIKE 20.  
FT DOMAIN 1197 1245 EGF-LIKE 21.  
FT DOMAIN 1247 1295 EGF-LIKE 22.  
FT DOMAIN 1298 1346 EGF-LIKE 23.  
FT DOMAIN 1352 1396 EGF-LIKE 24.  
FT DOMAIN 1397 1444 EGF-LIKE 25.  
FT DOMAIN 1446 1492 EGF-LIKE 26.  
FT DOMAIN 1494 1542 EGF-LIKE 27.  
FT DOMAIN 1544 1592 EGF-LIKE 28.  
FT DOMAIN 1594 1642 EGF-LIKE 29.

FT DOMAIN 1644 1692 EGF-LIKE 30.  
FT DOMAIN 1693 1739 EGF-LIKE 31.  
FT DOMAIN 1740 1788 EGF-LIKE 32.  
FT DOMAIN 1797 1847 EGF-LIKE 33.  
FT DOMAIN 1849 1900 EGF-LIKE 34.  
FT DOMAIN 1902 1950 EGF-LIKE 35.  
FT DOMAIN 1952 1998 EGF-LIKE 36.  
FT DOMAIN 2000 2047 EGF-LIKE 37.  
FT DOMAIN 2047 2095 EGF-LIKE 38.  
FT DOMAIN 2117 2156 EGF-LIKE 39.  
FT DOMAIN 2155 2284 SEA 1.  
FT DOMAIN 2293 2332 EGF-LIKE 40.  
FT DOMAIN 2331 2457 SEA 2.  
FT DOMAIN 2460 2504 EGF-LIKE 41.  
FT DOMAIN 2508 2556 EGF-LIKE 42.  
FT DOMAIN 2556 2608 EGF-LIKE 43.  
FT DOMAIN 2612 2657 EGF-LIKE 44, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 2657 2693 EGF-LIKE 45.  
FT DISULFID 2 13 BY SIMILARITY.  
FT DISULFID 7 22 BY SIMILARITY.  
FT DISULFID 24 43 BY SIMILARITY.  
FT DISULFID 50 63 BY SIMILARITY.  
FT DISULFID 57 72 BY SIMILARITY.  
FT DISULFID 74 93 BY SIMILARITY.  
FT DISULFID 104 120 BY SIMILARITY.  
FT DISULFID 112 129 BY SIMILARITY.  
FT DISULFID 131 143 BY SIMILARITY.  
FT DISULFID 149 163 BY SIMILARITY.  
FT DISULFID 157 172 BY SIMILARITY.  
FT DISULFID 174 191 BY SIMILARITY.  
FT DISULFID 198 214 BY SIMILARITY.  
FT DISULFID 208 223 BY SIMILARITY.  
FT DISULFID 225 244 BY SIMILARITY.  
FT DISULFID 251 265 BY SIMILARITY.  
FT DISULFID 259 274 BY SIMILARITY.  
FT DISULFID 276 295 BY SIMILARITY.  
FT DISULFID 317 330 BY SIMILARITY.  
FT DISULFID 324 339 BY SIMILARITY.  
FT DISULFID 341 353 BY SIMILARITY.  
FT DISULFID 359 371 BY SIMILARITY.  
FT DISULFID 365 380 BY SIMILARITY.  
FT DISULFID 382 401 BY SIMILARITY.  
FT DISULFID 408 421 BY SIMILARITY.  
FT DISULFID 415 430 BY SIMILARITY.  
FT DISULFID 432 451 BY SIMILARITY.  
FT DISULFID 458 472 BY SIMILARITY.  
FT DISULFID 466 481 BY SIMILARITY.  
FT DISULFID 483 502 BY SIMILARITY.  
FT DISULFID 709 725 BY SIMILARITY.  
FT DISULFID 717 734 BY SIMILARITY.  
FT DISULFID 736 749 BY SIMILARITY.  
FT DISULFID 754 768 BY SIMILARITY.  
FT DISULFID 762 778 BY SIMILARITY.  
FT DISULFID 780 793 BY SIMILARITY.  
FT DISULFID 809 822 BY SIMILARITY.  
FT DISULFID 816 831 BY SIMILARITY.  
FT DISULFID 833 845 BY SIMILARITY.  
FT DISULFID 851 867 BY SIMILARITY.  
FT DISULFID 859 876 BY SIMILARITY.  
FT DISULFID 878 891 BY SIMILARITY.  
FT DISULFID 896 909 BY SIMILARITY.  
FT DISULFID 903 918 BY SIMILARITY.  
FT DISULFID 920 939 BY SIMILARITY.  
FT DISULFID 946 959 BY SIMILARITY.  
FT DISULFID 953 968 BY SIMILARITY.  
FT DISULFID 970 989 BY SIMILARITY.  
FT DISULFID 996 1010 BY SIMILARITY.  
FT DISULFID 1004 1019 BY SIMILARITY.  
FT DISULFID 1021 1038 BY SIMILARITY.  
FT DISULFID 1047 1060 BY SIMILARITY.  
FT DISULFID 1054 1070 BY SIMILARITY.  
FT DISULFID 1072 1090 BY SIMILARITY.  
FT DISULFID 1097 1113 BY SIMILARITY.

Query	2	NPKNKCLQS--CNSERDSYRNOACHARNCLTKVKECEEEGI---PRPPRPQHP	53
Db	1746	DPRLNTGCRNMLCYDEPRGGR--CECRNGMDRSPDSQGRGVCPEPPPPPPPPRHR	1800
Query Match	Best Local Similarity	19.8%	Score 62.5; DB 1; Length 3051;
Matches	15; Conservative	10; Mismatches	24; Indels 9; Gaps 3;
RESULT	8		
EAR2_MOUSE			
ID	EAR2_MOUSE	STANDARD;	PRT; 390 AA.
AC	PA3136: 061504:		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-Oct-2000 (Rel. 40, Last annotation update)		
DE	ORF66 NUCLEAR RECEPTOR EAR-2 (V-ERBA RELATED PROTEIN EAR-2).		
DE	NR5F1 OR ERBA12 OR EAR2 OR EAR-2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		

```

RP [1] SEQUENCE FROM N.A.
RX MEDLINE=95034311; PubMed=7947324;
RA Jonk L.J.C., de Jonge M.J., Pals C.E.G.M., Missink S.,
RA Vervaeke J.M.A., Schoorlemmer J.V., Kruijer W.;
RT "Cloning and expression during development of three murine members of
RL Mech. Dev. 47:81-97(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PITUITARY;
RX MEDLINE=94252589; PubMed=8194772;
RA Barnhart K.M., Mellon P.L.;
RT "The sequence of a murine cDNA encoding Ear-2, a nuclear orphan
RL receptor.";
CC CC -1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC CC -1 SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC CC -N2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76654; CAAS4097.1; -.
DR EMBL; L25674; AAA37532.1; -.
DR HSSP; P19793; ZML.
DR MGD; MG1:99530; ERBAL2.
DR INTERPRO; IPR000536; -.
DR INTERPRO; IPR001628; -.
DR INTERPRO; IPR001723; -.
DR INTERPRO; IPR003068; -.
DR PFAM; PF00104; hormone-rec: 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR01282; COUPINFACOR.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KM Zinc-finger.
FT FT Dna_bind 57 122. C4-TYPE ZINC FINGERS (TWO).
FT FT Zn_fing 57 77 C4-TYPE.
FT FT Zn_fing 93 117 C4-TYPE.
FT FT Conflic 10 10 D-> G (IN REF. 2) .
FT FT Conflic 152 152 P-> H (IN REF. 2) .
FT FT Conflic 164 165 OL-> HV (IN REF. 2) .
FT FT Conflic 196 198 LLF-> AV (IN REF. 2) .
FT FT Conflic 239 241 LPL-> VAV (IN REF. 2) .
SQ SQ CONFICT 357 357 K-> Q (IN REF. 2) .
SQ SQ SEQUENCE 357 41982 MM; 4CZ7LF439CAF157F CRC64;

Query Match 19.7%; Score 62; DB 1; Length 390;
Best Local Similarity 33.3%; Pred. No. 8.3;
Matches 16; Conservative 6; Mismatches 18; Indels 8; Gaps 2;

QY 11 SCNSERP-----SYRNOACHAR--CNLLKVEKECEGELIPRRPPR 50
      :| | | :|| : | | : || : | | |
Db   92 TCGRSNRDQIDQHNRNCQCYCRALKCFVRGMKAEVQRIRPALPG 139

RESULT          9
EAR2_RAT        STANDARD;              PRT;       390 AA.
AC O09017;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ORPHAN NUCLEAR RECEPTOR EAR-2 (V-ERRA RELATED PROTEIN EAR-2)
```

```

DE (OVALBUMIN UPSTREAM PROMOTER GAMMA NUCLEAR RECEPTOR RCOUPG).
GN NR2FE OR ERBA12 OR EBR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PIUITARY;
RA Boutin J.M., Ronsin B., Devost D., Lipkin S.M., Rosenfeld M.G.,
RA Morel G.;
RT "Cloning and expression of COUP and COUPg in the rat anterior
RT pituitary.";
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RL -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF003926; AAB61296.1; -.
DR HSSP; P19793; 2NL.
DR INTERPRO: IPR000536; -.
DR INTERPRO: IPR001628; -.
DR INTERPRO: IPR001723; -.
DR PFAM; PF00104; hormone_rec; 1.
DR PFAM; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STROHORMONER.
DR PRINTS; PR01282; COUPNFACTOR.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KM Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KM Zinc-finger.
FT DNA_BIND 57 122 C4-TYPE ZINC FINGERS (TWO).
FT ZN_FING 57 77 C4-TYPE.
FT ZN_FING 93 117 C4-TYPE.
SQ SEQUENCE 390 AA: 41729 MW: B08FAC93F76A7783 CRC64;

```

```

Query Match 19.0%; Score 60; DB 1; Length 390;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 16; Conservative 6; Mismatches 18; Indels 8; Gaps 2;
QY 11 SCNSEPD-----SYRQACGAR---CNLKVKEKECECEGPIRRPRRP 50
DB 92 TCRSNRDCCQIDQHHNRNCCQCYRLKCFRVMGRKREAVQGPPIPHALGPR 139

```

```

RESULT 10
CCAD_RAT STANDARD: PRT: 2203 AA.
AC P27732; Q63491; Q63492; Q62691; Q01542; Q09022; Q09023; Q09024;
AC Q62815;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1D SUBUNIT (CALCIUM
DE CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE, ISOFORM 2) (RAT BRAIN CLASS D)
DE (RBD).
GN CACNA1D OR CACNA1A2 OR CCH1A2 OR CACH3 OR CACNA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS CACNA4 AND CACNA5).
RC TISSUE=INSULINOMA.
RX MEDLINE=95280950; PubMed=7760845;
RA Ihara Y., Yamada Y., Fujii Y., Gonoji T., Yano H., Yasuda K.,

```

```

RA Inagaki N., Saino Y., Saino S.;
RT "Molecular diversity and functional characterization of voltage-
RT dependent calcium channels (CACNA4) expressed in pancreatic beta-
RT cells.";
RL Mol. Endocrinol. 9:121-130(1995).
RN [2]
RP SEQUENCE FROM N.A. (TRUNCATED ISOFORM AND CLONES RB9/11/34/48).
RC TISSUE=BRAIN;
RX MEDLINE=91299338; PubMed=1648940;
RA Hul A., Ellinor P.T., Krizanov O., Wang J.-J., Diebold R.J.,
RA Schwartz A.;
RT "Molecular cloning of multiple subtypes of a novel rat brain isoform
RT of the alpha-1 subunit of the voltage-dependent calcium channel.";
RL Neuron 7:35-44(1991).
RN [3]
RP SEQUENCE OF 1-125 FROM N.A.
RX MEDLINE=96040125; PubMed=7553731;
RA Kamp T.J., Mitas M., Fields K.L., Asch S., Chin H., Marban E.,
RA Nirenberg M.;
RT "Transcriptional regulation of the neuronal L-type calcium channel
RT alpha 1D subunit gene.";
RL Cell. Mol. Neurobiol. 15:307-326(1995).
RN [4]
RP SEQUENCE OF 1100-1410 FROM N.A. (CLONES RKC5 AND RKC6).
RC TISSUE=KIDNEY;
RX MEDLINE=93066265; PubMed=1279681;
RA Yu A.S.L., Hebert S.C., Brenner B.M., Lytton J.;
RT "Molecular characterization and nephron distribution of a family of
RT transcripts encoding the pore-forming subunit of Ca2+ channels in the
RT kidney.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10494-10498(1992).
RN [5]
RP SEQUENCE OF 1218-1498 FROM N.A. (ISOFORM ROB3).
RC TISSUE=OSTEOSARCOMA;
RX MEDLINE=96074617; PubMed=7479909;
RA Barry E.L.R., Gesek P.A., Froehner S.C., Friedman P.A.;
RT "Multiple calcium channel transcripts in rat osteosarcoma cells:
RT selective activation of alpha 1D isoform by parathyroid hormone.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10914-10918(1995).
RN [6]
RP SEQUENCE OF 1200-1493 FROM N.A. (RH1; RH2; CACNA4/B & DELTA-1VS3/54).
RC TISSUE=HEPATOMA;
RX MEDLINE=97376179; PubMed=9232351;
RA Breton H.M., Harland M.L., Frosco M., Pelronjic T.,
RA Barritt G.J.;
RT "Novel variants of voltage-operated calcium channel alpha-1 subunit
RT transcripts in a rat liver-derived cell line: deletion in the IVS4
RT voltage sensing region.";
RL Cell Calcium 22:35-52(1997).
RN [7]
RP SEQUENCE OF 1307-1479 FROM N.A. (CLONE RBD-55).
RX MEDLINE=90239020; PubMed=1692134;
RA Snutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.;
RT "Rat brain expresses a heterogeneous family of calcium channels.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3391-3395(1990).
RN [8]
RP -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
RP ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
RP IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
RP CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
RP CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1D
RP GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)
RP CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)
RP GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP),
RP PHENYLALKYLAMINES, BENZOTHIAZEPINES, AND BY OMEGA-AGATOXIN-11IA
RP (OMEGA-AGA-11IA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-
RP GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
RN [9]
RP -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
RP COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
RP IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
RP FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
RP SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
RP CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
RP LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.

```

```

CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- ALTERNATIVE PRODUCTS: MULTIPLE ISOFORMS ARE CREATED BY ALTERNATIVE
CC SPLICING EVENTS, WHICH SEEM TO OCCUR IN A TISSUE-SPECIFIC MANNER.
CC THE SEQUENCE SHOWN HERE IS THAT OF CACNA4. THE REGION SEQUENCED IN
CC ROB3 AND RKC5 IS IDENTICAL TO CACNA4.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, PANCREATIC ISLETS AND B-
CC LYMPHOCYTES.
CC -I- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -I- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D38101: BAA07282.1; -
DR EMBL: D38102: BAA07283.1; -
DR EMBL: M57682: AAA42015.1; -
DR EMBL: U14005: AAB60515.1; -
DR EMBL: M99221: AAA40895.1; -
DR EMBL: U31772: AAA89156.1; -
DR EMBL: U49126: AAB61634.1; -
DR EMBL: U49127: AAB61635.1; -
DR EMBL: U49128: AAB61636.1; -
DR INTERPRO: IPR000636; -
DR INTERPRO: IPR002077; -
DR INTERPRO: IPR003091; -
DR PFAM: PF00520: Ion.trans. 4.
DR PRINTS: PR00167; CACHANNEL.
DR PRINTS: PR00169; KCHANNEL.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Alternative splicing.
FT REPEAT 112 408
FT REPEAT 528 774
FT REPEAT 892 1174
FT REPEAT 1211 1486
FT DOMAIN 1 126
FT TRANSMEM 127 145
FT TRANSMEM 146 163
FT TRANSMEM 164 183
FT TRANSMEM 184 195
FT TRANSMEM 196 214
FT TRANSMEM 215 235
FT TRANSMEM 236 254
FT TRANSMEM 255 273
FT TRANSMEM 274 293
FT TRANSMEM 294 381
FT TRANSMEM 382 406
FT TRANSMEM 407 582
FT TRANSMEM 583 602
FT TRANSMEM 603 617
FT TRANSMEM 618 636
FT TRANSMEM 637 644
FT TRANSMEM 645 663
FT TRANSMEM 664 673
FT TRANSMEM 674 692
FT TRANSMEM 693 711
FT TRANSMEM 712 732
FT TRANSMEM 733 786
FT TRANSMEM 787 811
FT TRANSMEM 812 945
FT TRANSMEM 946 980
FT DOMAIN 981 1000

```

FT	DOMAIN	1001	1012	CYTOPLASMIC (POTENTIAL).
FT	TRANSNM	1031	1031	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1032	1037	EXTRACELLULAR (POTENTIAL).
FT	TRANSNM	1038	1057	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1058	1076	CYTOPLASMIC (POTENTIAL).
FT	TRANSNM	1077	1096	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1097	1186	EXTRACELLULAR (POTENTIAL).
FT	TRANSNM	1187	1207	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1208	1264	CYTOPLASMIC (POTENTIAL).
FT	TRANSNM	1265	1283	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1284	1298	EXTRACELLULAR (POTENTIAL).
FT	TRANSNM	1299	1318	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1319	1325	CYTOPLASMIC (POTENTIAL).
FT	TRANSNM	1326	1347	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1348	1357	EXTRACELLULAR (POTENTIAL).
FT	TRANSNM	1358	1377	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1378	1396	CYTOPLASMIC (POTENTIAL).
FT	TRANSNM	1397	1416	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1417	1483	EXTRACELLULAR (POTENTIAL).
FT	TRANSNM	1484	1508	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1509	2203	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	1	7	POLY-MET.
FT	DOMAIN	886	712	POLY-LEU.
FT	DOMAIN	897	897	POLY-GLU.
FT	DOMAIN	429	446	BINDING TO THE BETA SUBUNIT (BY SIMILARITY).
FT	SITE	364	364	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	763	763	(BY SIMILARITY).
FT	SITE	1160	1160	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	1450	1450	(BY SIMILARITY).
FT	SITE	1450	1450	CALCIUM ION SELECTIVITY AND PERMEABILITY
FT	SITE	1450	1450	(BY SIMILARITY).
FT	BINDING	1134	1224	TO DIHYDROPRIDINES (BY SIMILARITY).
FT	BINDING	1464	1330	TO DIHYDROPRIDINES (BY SIMILARITY).

Query Match 19.0%; Score 60; DB 1; Length 2203;  
Best Local Similarity 34.9%; Pred. No. 65;  
Matches 22; Conservative 9; Mismatches 18; Indels 14; Gaps 4.

Oy	1	ENPKRINK---CIQSCSER---DSTRNOA---CHARCNLKVKECEGEGRPR---P	46
Db	843	ENRKNKKEVNOIANSNDKVTIDYQGEAEKDKPPCDVPVGEFEFEFEDEPEVPAGP	902
Oy	47	RPR	49
Db	903	RPR	905

RESULT 11  
BARI\_CHITE  
ID BARI\_CHITE STANDARD: PRT: 174 AA.  
AC P02849;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE BALBANI RING PROTEIN 1 (GIANT SECRETORY PROTEIN 1-A) (GSP-1A) (FRAGMENT).  
GN Chironomus tentans (Midge).  
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
OC Chironomidae; Chironomidae; Chironomidae; Chironomus.  
RN [1]  
RP SEQUENCE OF 1-116 FROM N.A. (CLONE PCTBRI-1).  
RC TISSUE=Salivary Gland;  
RX MEDLINE=83238361; PubMed=6305953;  
RA Case S.T., Byers M.R.;  
RT "Repeated nucleotide sequence arrays in Balbiani ring 1 of Chironomus tentans contain internally nonrepeating and subrepeating elements.";  
J. Biol. Chem. 258:7793-7799(1983).  
LN [2]

```

RP SEQUENCE OF 7-174 FROM N.A. (CLONE PCT21).
RC TISSUE-SALIVARY GLAND;
RA Wieslander L., Sumegi J., Daneholt B.;
RT "Evidence for a common ancestor sequence for the Balbiani ring 1 and
RL Balbiani ring 2 genes in Chironomus tentans.";
RC Proc. Natl. Acad. Sci. U.S.A. 79:6956-6960(1982).
CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE, THE LARVAL TUBE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K00447; AAA28238.1; -
DR EMBL; J01055; AAA28236.1; -
DR PIR; A03339; BQICIT.
KW Repeat.
FT FT 1 1 4 X 11 AA TANDEM REPEATS.
FT REPEAT 42 85 1-1.
FT REPEAT 53 63 1-2.
FT REPEAT 64 74 1-3.
FT REPEAT 75 85 1-4.
FT DOMAIN 124 167 4 X 11 AA TANDEM REPEATS.
FT REPEAT 124 134 2-1.
FT REPEAT 135 145 2-2.
FT REPEAT 146 156 2-3.
FT REPEAT 157 167 2-4.
FT VARIANT 40 40 N -> K (IN CLONE PCT21).
FT VARIANT 57 57 G -> E (IN CLONE PCT21).
FT VARIANT 72 72 K -> R (IN CLONE PCT21).
FT VARIANT 86 86 R -> G (IN CLONE PCT21).
FT NON_TER 174 174
FT SEQUENCE 174 AA; 18920 MW; E86E616C6413D751 CRC64;

Query Match 18.9%; Score 59.5; DB 1; Length 174;
Best Local Similarity 26.5%; Pred. No. 7.4;
Matches 18; Conservative 11; Mismatches 24; Indels 15; Gaps 4;

QY 3 PKHNLC--LQSCNSERSYSRN---QACHARCNLL--KVEKECEGEGLPRP-----R 47
DB 82 PRPEKSGSAMKRAEKARNGRGNASKRCCTSGKRSKRSKSPKSGSKPRPEKRSKSK 141
QY 48 PRPOHPR 55
DB 142 PRPEKPSK 149

RESULT 12
MS11_DROHY STANDARD; PRT; 344 AA.
AC 008695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AXONEME-ASSOCIATED PROTEIN MS1101(1).
GN MS1101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridiidae; Drosophilidae; Drosophila.
RN 11
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-TESTIS;
RX MEDLINE-94200512; PubMed-8150205;
RA Neesen J., Buemann H., Heinlein U.A.;

```

RT "The *Drosophila* hydei gene Dmhist10(1) encodes a testis-specific, repetitive, axoneme-associated protein with differential abundance in y chromosomal deletion mutant files."

RL dev. Biol. 162:414-425(1994).

CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS ASSOCIATED WITH AXONEMAL STRUCTURES.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND SPERMATID BUNDLES.

CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.

CC -1- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-slb.ch/announce/> or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch))

CC -----

DR EMBL, X73480; CAA51875.1; -.

DR PIR, S34153; S34153.

DR HSSP: P01032; 1C5A.

DR FLXBASE: FB90q011816; Dbydymst101(1).

KW Sperm: Repeat: Multigene family; Polymorphism.

FT DOMAIN 58 337 19 X 16 AA APPROXIMATE TANDEM REPEATS OF K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.

FT -----

SO SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;

QY 1 ENPKHNKCIQSCNSRSDSYRNQACIAPCNLAK--VEKEECE 40  
I : I I : : I : : : I : I : I : I I I I I

Db 191 EAAEKKKAEEAKKKEKEAAEKKKCEERAKKKEEAKKKEE 232

RESULT 13

TS44\_GIALA

ID TS44\_GIALA STANDARD; PRT: 713 AA.

AC P21849;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE MAJOR SURFACE-LABELLED TROPHOZOITE ANTI GEN PRECURSOR.

GN TSA 417.

OS *Giardia lamblia* (*Giardia intestinalis*).

OC Eukaryota: Diplomonadida; Hexamitidae; *Giardinae*; *Giardia*.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 30957 / WB:

RC MEDLINE=90280395; PubMed=2352292;

RA Gillin F.D., Hagblom P., Harwood J., Aley S.B., Reiner D.S., McCallery M., So M., Giney D.G.;

RT "Isolation and expression of the gene for a major surface protein of *Giardia lamblia*."

RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).

RN [2]

RP SEQUENCE OF 480-620 FROM N.A.

RC STRAIN=AD-1;

RC MEDLINE=93314970; PubMed=8325510;

RA Ey P.L., Mayrhofer G.;

RT "Two genes encoding homologous 70-kDa surface proteins are present within individual trophozoites of the binucleate protozoan parasite *Giardia intestinalis*."

RT Gene 129:257-262(1993).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- DOMAIN: CONTAINS 29 REPEATS OF THE CXKX MOTIF.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -





CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U62771; AAC61296.1; -.  
DR INTERPRO: IPR000276; -.  
DR PFAM: PF00001; 7tm\_1; 2.  
DR PRINTS: PR00237; GPCRHHODOPSN.  
DR PROSITE: PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.  
DR PROSITE: PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
FT DOMAIN 1 28 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 29 53 POTENTIAL.  
FT DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 65 87 POTENTIAL.  
FT DOMAIN 88 102 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 103 124 POTENTIAL.  
FT DOMAIN 125 147 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 148 167 POTENTIAL.  
FT DOMAIN 168 239 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 240 259 POTENTIAL.  
FT DOMAIN 260 520 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 521 545 POTENTIAL.  
FT DOMAIN 546 551 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 552 575 POTENTIAL.  
FT DOMAIN 576 638 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 101 230 BY SIMILARITY.  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 638 AA: 70000 MW: 65FA928B5C01D34F CRC64;

Query Match 18.6%; Score 58.5; DB 1; Length 638;  
Best Local Similarity 29.6%; Pred. No. 30;  
Matches 16; Conservative 8; Mismatches 17; Indels 13; Gaps 3;

QY 10 QSCNSERDSYRN-----QACHARNLLKYEKECEGEI-PPRPPQPH 52  
: | | | | : | | | | : | | | | : |  
Db 402 RSVGGERSEFENSTRNGRSTRAKLCGRG--LAIEFDICSSGCSPTKRIKEH 453

Search completed: March 1, 2001, 16:20:36  
Job time: 211 sec